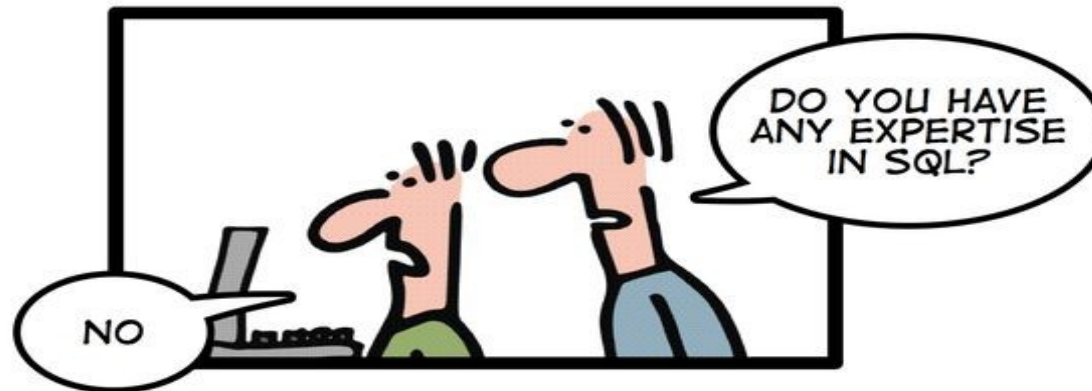


Bases de datos NoSQL



Hugo González
Noviembre 17, 2011

HOW TO WRITE A CV



Leverage the NoSQL boom

¿Qué es una Base de Datos?



Una base de datos o banco de datos es un **conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso**. En este sentido, una biblioteca puede considerarse una base de datos compuesta en su mayoría por documentos y textos impresos en papel e indexados para su consulta. Actualmente, la mayoría de las bases de datos están en formato digital (electrónico), que ofrece un amplio rango de soluciones al problema de almacenar datos.

Gestor de Base de Datos

Los sistemas de gestión de bases de datos (en inglés database management system, abreviado DBMS) son un tipo de software muy específico, dedicado a servir de interfaz entre la base de datos, el usuario y las aplicaciones que la utilizan.

El propósito general de los sistemas de gestión de bases de datos es el de manejar de manera clara, sencilla y ordenada un conjunto de datos que posteriormente se convertirán en información relevante para una organización.

Sistemas disponibles

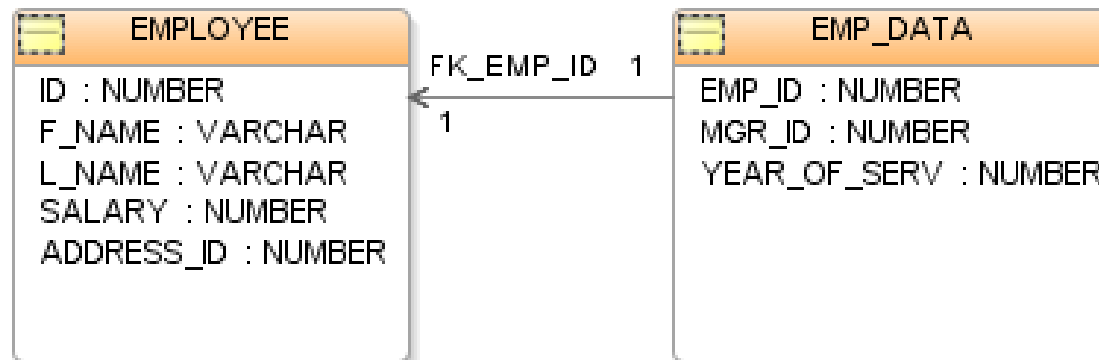
- PostgreSQL (<http://www.postgresql.org> Postgresql) Licencia BSD
- Firebird basada en la versión 6 de InterBase, Initial Developer's PUBLIC LICENSE Version 1.0.
- SQLite (<http://www.sqlite.org> SQLite) Licencia Dominio Público
- DB2 Express-C (<http://www.ibm.com/software/data/db2/express/>)
- Apache Derby (<http://db.apache.org/derby/>)
- MariaDB (<http://mariadb.org/>)
- MySQL (<http://dev.mysql.com/>)
-
- Oracle, MSSQL, Access, Informix ...

Tipos de Bases de Datos

- Bases de datos jerárquicas
- Base de datos de red
- Bases de datos transaccionales
- **Bases de datos relacionales**
- Bases de datos multidimensionales
- Bases de datos orientadas a objetos
- Bases de datos documentales
- Bases de datos deductivas

Bases de datos relacionales

- Una base de datos relacional es una base de datos que cumple con el **modelo relacional**, el cual es el modelo más utilizado en la actualidad para implementar bases de datos ya planificadas. Permiten establecer interconexiones (**relaciones**) entre los datos (que están guardados en tablas), y a través de dichas conexiones relacionar los datos de ambas tablas, de ahí proviene su nombre: "Modelo Relacional". Tras ser postuladas sus bases en 1970 por Edgar Frank Codd, de los laboratorios IBM en San José (California), no tardó en consolidarse como un nuevo paradigma en los modelos de base de datos.¹



SQL

- El lenguaje de **consulta estructurado** o SQL (por sus siglas en inglés structured query language) es un lenguaje declarativo de **acceso a bases de datos relacionales** que permite especificar diversos tipos de operaciones en estas. Una de sus características es el manejo del **álgebra y el cálculo relacional** permitiendo efectuar consultas con el fin de recuperar -de una forma sencilla- información de interés de una base de datos, así como también hacer cambios sobre ella.

NoSQL

- NoSQL es un término usado en para agrupar una serie de almacenes de datos no relacionales que no proporcionan garantías ACID. Normalmente no tienen esquemas fijos de tablas ni sentencias "join".

El término NOSQL fue acuñado a principios de 1999 por un empleado de Rackspace, Eric Evans, cuando Johan Oskarsson de Last.fm quiso organizar un evento para discutir bases de datos distribuidas de código abierto. El nombre intenta describir el surgimiento de un número creciente de bases de datos no relacionales y distribuidas que no suelen proveer garantías ACID. El término ACID hace referencia a un conjunto de características necesarias para que una serie de instrucciones puedan ser consideradas como una transacción.

ACID

- Atomicidad: es la propiedad que asegura que la operación se ha realizado o no, y por lo tanto ante un fallo del sistema no puede quedar a medias.
- Consistencia: Integridad. Es la propiedad que asegura que sólo se empieza aquello que se puede acabar. Por lo tanto se ejecutan aquellas operaciones que no van a romper las reglas y directrices de integridad de la base de datos.
- Aislamiento: es la propiedad que asegura que una operación no puede afectar a otras. Esto asegura que la realización de dos transacciones sobre la misma información sean independientes y no generen ningún tipo de error.
- Durabilidad: es la propiedad que asegura que una vez realizada la operación, ésta persistirá y no se podrá deshacer aunque falle el sistema.

Bases de datos documentales

- CouchDB, de Apache CouchDB
- MongoDB, de 10gen MongoDB
- RavenDB, de Hibernating Rhinos. [1]
- BaseX
- eXist
- SimpleDB
- IBM Lotus Domino
- Terrastore

Bases de datos en grafo

- Neo4j
- DEX
- AllegroGraph
- OrientDB
- InfiniteGraph
- Sones GraphDB
- InfoGrid
- HyperGraphDB

Bases de datos clave/valor

- Cassandra, de Apache The Apache Cassandra
- BigTable, de Google
- Dynamo, de Amazon
- Project Voldemort, de LinkedIn
- Riak

Bases de datos multivalor

- OpenQM
- Extensible storage engine

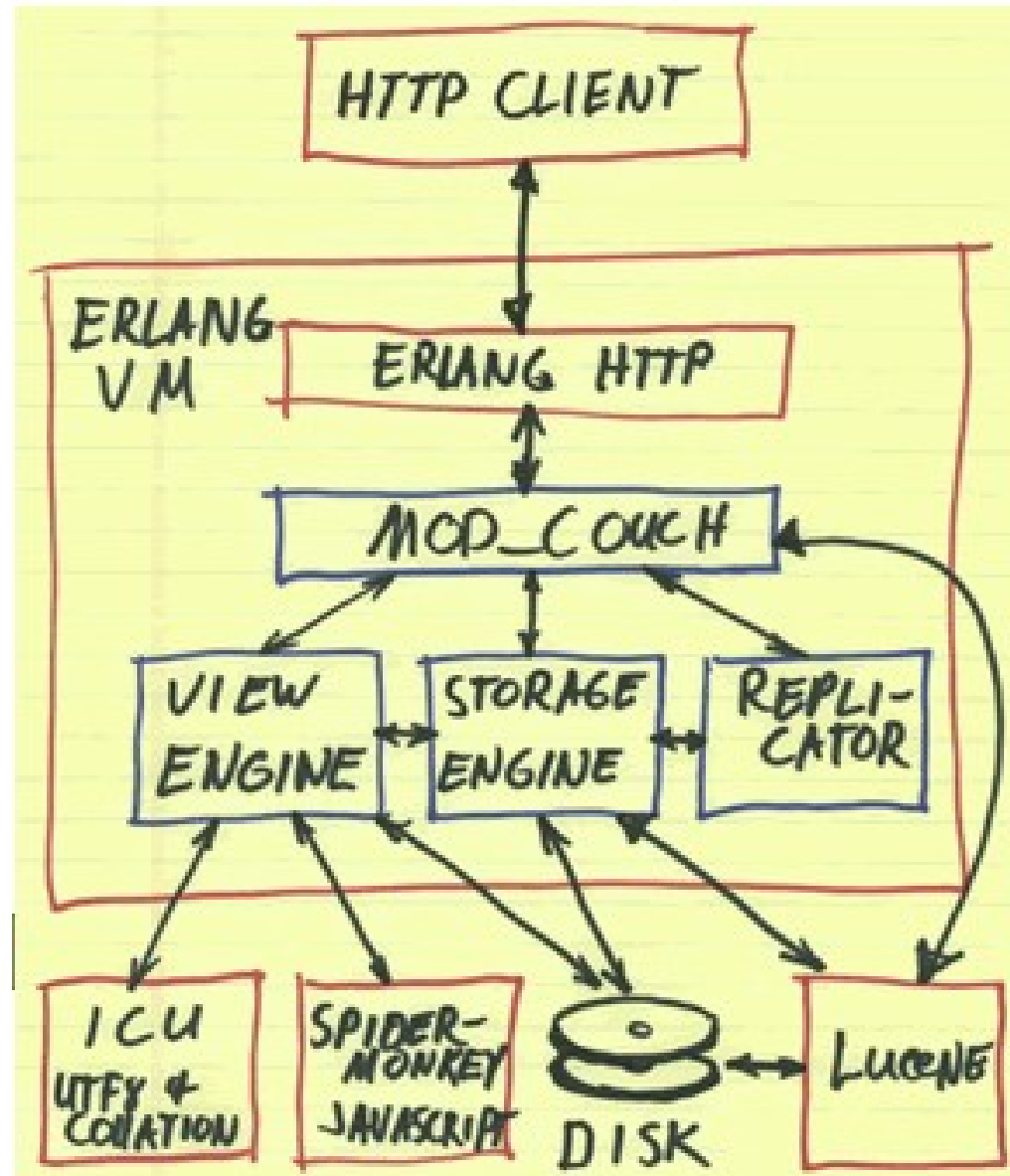
Bases de datos orientadas a objetos

- Zope Object Database
- db4o
- GemStone S
- Objectivity/DB

Bases de datos tabular

- HBase, de Apache
- BigTable, de Google
- Hypertable

CouchDB - Relax



Algunas características

- Servicios web RESTful (se puede usar cualquier lenguaje para interactuar)
- Habla JSON
- Estructuras variables / flexibles
- Uso de Javascript para funciones de map-reduce para obtener consultas
- BD distribuida
- Manejo de historial de documentos

Algunas desventajas

- Cambio de modelo / paradigma
- Uso de esquemas diferentes.

Map-reduce

```
void map(String name, String document):  
    // name: document name  
    // document: document contents  
    for each word w in document:  
        EmitIntermediate(w, "1");  
  
void reduce(String word, Iterator partialCounts):  
    // word: a word  
    // partialCounts: a list of aggregated partial counts  
    int sum = 0;  
    for each pc in partialCounts:  
        sum += ParseInt(pc);  
    Emit(word, AsString(sum));
```

Ejemplo !

Preguntas